

Universidade Federal de Pernambuco  
1ª Avaliação de Análise Exploratória de Dados - 09/08/2022

Professor: Dr. Vinícius Quintas Souto Maior

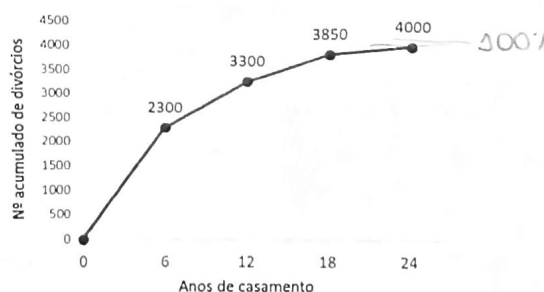
Atenção: As respostas de todas as questões devem estar acompanhadas dos cálculos.

- (1) Para melhor entender a situação no Brasil, os dados do portal <https://cidades.ibge.gov.br/brasil/panorama> (Síntese de Indicadores Sociais 2018) fornecem a *média salarial* nas 27 unidades federativas da união. A frequência de ocorrência considerando 5 classes encontra-se mostrada na tabela a seguir. As três UF's de pior média salarial eram AL, MA e PA. Os três de maior média salarial eram DF, SP e AP.

Média salarial (R\$)	Frequência
1.700   - 2.000	3
2.000   - 2.300	12
2.300   - 2.600	6
2.600   - 2.900	3
2.900   - 3.200	3

Separando-se por região e tomando as respectivas médias salariais, chega-se aos valores R\$ 2.451,00, R\$ 2.070,00, R\$ 2.516,00, R\$ 2.495,00 e R\$ 2.881,00 para, respectivamente, as regiões Norte, Nordeste, Sudeste, Sul e Centro Oeste.

- a) (1,25) Esboce o histograma das médias salariais no Brasil (ano referência 2018).  
b) (1,25) Trace o box-plot que corresponde aos dados agrupados por região do país. Quais seus comentários?
- (2) Dada a distribuição de frequências acumuladas do número de divórcios segundo a duração do casamento, representada no polígono de frequências acumuladas a seguir, pode-se:



- a) (1,0) Reproduza as informações do gráfico em uma tabela de distribuição de frequências para o tempo de duração do casamento.  
b) (1,0) 25% dos divórcios se deram com casamentos de até quantos anos?

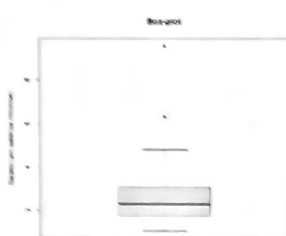
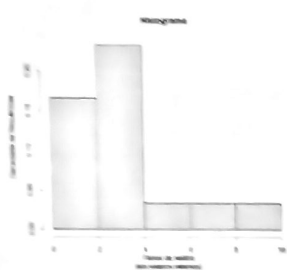
2.5  
x

sexo e lugar

(3) Considerando o trecho do relatório a seguir:

"Assim sendo, podemos concluir que este banco, em 2021, contou com a colaboração de 345 profissionais, distribuídos pelas nossas 5 agências, a saber: Niterói, 43; Rio de Janeiro, 102; São Paulo, 98; Belo Horizonte, 75; Vitória, 27. Em Niterói, 38 eram do sexo masculino e no Rio de Janeiro, 87. Apenas em Vitória não existiam funcionárias, mas em São Paulo trabalharam 11 delas, enquanto que em Belo Horizonte, apenas 3."

- a)(1.0) Identifique as variáveis em estudo, classificando-as. Justifique a sua resposta.
- b)(1.5) Represente tabularmente (distribuição de frequências) e graficamente os dados descritos no relatório para cada uma das variáveis em estudo.
- (4) Diversas técnicas estatísticas são baseadas na suposição de que os dados provêm de uma distribuição normal ou, pelo menos, aproximadamente simétrica. Os dados referentes aos salários (em salários mínimos) dos funcionários da empresa X são: 1,07; 1,15; 1,43; 1,48; 1,99; 2,18; 2,24; 2,31; 2,31; 2,40; 2,84; 3,36; 4,81; 6,33; 9,54.



14/4 = 3.5  
1Q = 1.73  
2Q = 2.31  
3Q = 3.2

- a)(1.5) Apresente o gráfico de simetria, calcule alguma medida de assimetria e classifique a curva como aproximadamente simétrica, assimétrica positiva ou assimétrica negativa.
- b)(1.0) Calcule alguma medida de curtose e classifique a curva como mesocúrtica, leptocúrtica ou platocúrtica.
- (5) Com o objetivo de obter simetria, os dados da Questão 4 foram transformados como  $\ln(X)$  e  $X^{-0.7}$  e são apresentados na tabela abaixo juntamente com um breve resumo descritivo.

$\ln(X)$	0.07	0.14	0.36	0.39	0.69	0.78	0.81	0.84	0.84	0.87	1.04	1.21	1.57	1.84	2.25
$X^{-0.7}$	0.95	0.91	0.78	0.76	0.62	0.58	0.57	0.56	0.56	0.54	0.48	0.43	0.33	0.27	0.21

- c)(1.0) Qual das transformações obteve dados mais homogêneos? Comente.
- d)(1.0) Avalie qual das duas transformações será preferida para obtenção da simetria? Explique!

	Q(0.25)	Q(0.50)	Q(0.75)	Média	Var	Mín	Máx
$\ln(X)$	0.54	0.84	1.13	0.07	0.91	0.371	2.25
$X^{-0.7}$	0.45	0.56	0.69	0.21	0.57	0.046	0.95

nd